Develop Predictive Models Using Your Business Expertise

Clementine is an integrated data mining workbench, popular worldwide with data miners and business analysts alike. It supports the entire data mining process, enabling users to quickly access and prepare data for modeling, rapidly build and compare models, and then deploy models into real-time operational environments. And it seamlessly integrates data mining results with other business systems and processes, giving decision makers a clearer view of current conditions and insight into future events.

Clementine's highly visual user interface makes "train of thought" analysis possible. Clementine produces "streams," a visual map of your data mining steps, to automatically document the analytical process. By interacting with a stream, data miners can add business knowledge—the key to successful data mining—to the data mining process. The Clementine interface makes it easy to explore ideas or intuitive hunches. And because it allows you to focus on knowledge discovery, rather than on performing technical tasks such as writing code, you achieve results quickly.

Select from an unparalleled breadth of techniques

Clementine offers a full range of machine-learning and statistical techniques for clustering, classification, association, and prediction. SPSS Inc. has more than 35 years' experience in predictive analytics, and its algorithms are calibrated and verified to support the creation of powerful data mining models. With Clementine, you can be assured of having the analytical techniques you need to attain the most reliable results. Clementine 8.5 incorporates CARMA, a high-speed association detection algorithm, and offers a new association viewer.

Integrate with existing information systems

With Clementine, there's good news for your IT staff. Clementine is an open, standards-based solution that easily integrates with your organization's existing operational systems. In addition, Clementine leverages your database investment by performing much of the data preprocessing tasks—typically 80-90 percent of your data mining effort—within the database. Because preprocessing occurs on the database tier, rather than on the server or client tiers, you save time, maximize resources, and minimize network traffic. And, with Clementine 8.5, organizations using Microsoft® SQL Server™ 2000 Analysis Services can build Microsoft decision trees directly from the Clementine interface.

Put results to work in your organization

SPSS offers several deployment options to meet your needs for in-database or real-time scoring. These include saving models in industry-standard Predictive Model Markup Language (PMML) and embedding them as stand-alone files in other systems and applications. You can also export not only models but also all data mining steps—including data access, modeling, and post-processing—through Clementine Solution Publisher. These and other deployment options increase the usability of Clementine models and save your organization time and money.

With Clementine 8.5, companies can automatically deploy models built in Clementine into IBM® DB2® Data Warehouse Enterprise Edition. This release also enables companies to author and deploy streams from Clementine to the PredictiveCallCenter™ application from SPSS.

Clementine 8.5 offers enhanced performance and integration

New features provide additional power and integration, enabling your organization to further leverage your investment in information technology.

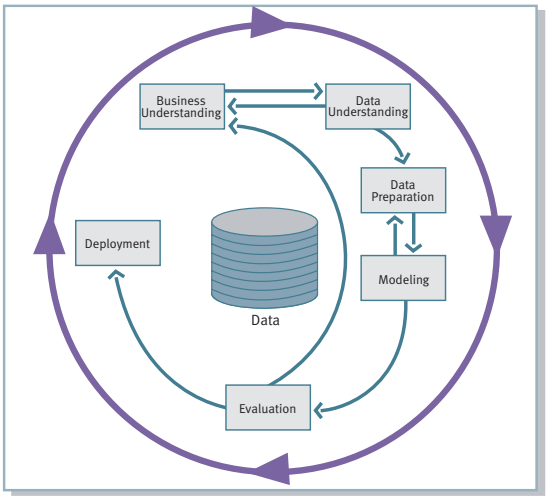
- If you use Microsoft SQL Server 2000 Analysis Services, you can use Clementine to build Microsoft decision trees
- If you use Clementine on the Sun℠ Solaris™ platform, you will enjoy improved performance in processing large datasets, because Clementine 8.5 is available as a 64-bit application on SPARC servers
- If you use Clementine with IBM DB2 databases, you can browse association models, and automatically deploy models built in Clementine into DB2 Data Warehouse Enterprise Edition
- If you use PredictiveCallCenter from SPSS, you can author and automatically deploy Clementine streams to PredictiveCallCenter's Interaction Builder component

Other new features include a new association viewer and CARMA, a high-speed association detection algorithm that delivers faster processing and gleans more detailed information from your data.

Features*

CRISP-DM

Help your company's analysts and data miners focus on business problem solving, rather than on the data mining technology. Clementine supports the CRoss-Industry Standard Process for Data Mining (CRISP-DM), the de facto industry standard, at every step of the process. Clementine projects can be efficiently organized using the CRISP-DM project manager.



The CRISP–DM process, as shown above, enables data miners to efficiently implement data mining projects that yield measurable business results.

Business understanding

Clementine's visual interface makes it easy for your data miners to apply business knowledge to data mining projects. In addition, optional, business-specific Clementine Application Templates (CATs) are available accelerate the attainment of data mining results. CATs ship with sample data that can be installed as flat files or as tables in a relational database schema. You may choose from the:

- CRM CAT**
- Telco CAT**
- Fraud CAT**
- Microarray CAT**
- Web mining CAT** (requires the purchase of Web Mining for Clementine)

Data understanding

- You can obtain a comprehensive "first look" at your data through Clementine's data audit node. Create summary statistics, histograms, and distribution graphics for each data field and display them in an easy-to-read matrix. Clementine also enables you to visually interact with output by generating subsets of data from graphs and tables.
- You can also:
  - Select a region of a graphic and view the selected information in a table, or use this information downstream
  - Create histograms, distributions, line plots, and point plots
  - Use Web association detection
  - Display 3-D, panel, and animated graphs
  - View data quickly through graphs, summary statistics, or an assessment of data quality

Data preparation

- Access data
  - Structured (tabular) data
    - Manage data security through the selected database and/or operating system
    - Access ODBC-compliant data sources with the SPSS Data Access Pack, which ships with Clementine. Drivers in this middleware pack include support for Oracle,® Microsoft SQL Server, IBM DB2, Informix,® and Sybase® databases.
    - Import delimited and fixed-width text, as well as any SPSS® and SAS® 6, 7, 8, and 9 file
  - Unstructured (textual) data
    - Automatically extract concepts from documents and from text notes in databases, using Text Mining for Clementine**
  - Web site data
    - Automatically extract Web site events from Web logs, using Web Mining for Clementine**
  - Data output
    - Work with delimited and fixed-width text, ODBC, Microsoft Excel, SPSS, and SAS 6, 7, 8, and 9 files
- Choose from various data-cleaning options
  - Remove or replace invalid data
  - Automatically fill in missing values
- Manipulate data
  - Work with complete record and field operations, including:
    - Field filtering, naming, derivation, and value replacement
    - Record selection, sampling, merging, and concatenation; sorting, aggregation and balancing; deriving new fields based on conditional criteria; and calculating new fields
    - Specialized manipulations for showing the "history" of values and converting set variables into flag variables
  - Automatically bin data based on n-tiles, mean and standard deviation, rank, or a fixed width
  - Re-group or collapse categories and transform one set of discrete values into another
  - Change the order of fields
- Merge records using any of the following types of joins:
  - Inner join
  - Full outer join
  - Partial outer join, both left and right
  - Anti-join (the opposite of an inner join)

** Features are subject to change based on the final product release.* ☐ *Symbol indicates a new feature.* ***Separately priced modules*

Modeling

- Prediction and classification
  - Neural networks (multi-layer perceptrons using error back-propagation, radial basis function, and Kohonen networks)
    - Browse the importance of the predictors
  - Decision trees and rule induction [C5.0 and Classification and Regression Trees (C&RT)]
    - Browse and interact with decision trees
    - Browse, collapse, and expand decision rules
  - Linear regression, logistic regression, and multinomial logistic regression
    - View model equation and advanced statistical output
- Clustering and segmentation
  - Kohonen network, K-means, and TwoStep
    - View cluster characteristics with a graphical viewer
- Association detection
  - GRI, Apriori, sequence
  - CARMA
  - Filter, sort, and subset association models with the new association model viewer
  - Use Apriori-, sequence-, and CARMA-generated models for scoring
- Data reduction
  - Factor analysis and principal components analysis
    - View model equation and advanced statistical output
- Combine models for greater accuracy through meta-modeling
  - Multiple models can be combined, or one model can be used directly for building a second model
- Import PMML-generated models created in other tools, such as AnswerTree® and SPSS for Windows®
- Clementine External Module Interface (CEMI) for custom algorithms
  - Add data access, modeling, and graphics
  - Create node palettes and generated model palettes, including PMML-generated models created in other tools, using custom CEMI icons
- An algorithm user manual is included, which explains the theories and methods behind the algorithms included in Clementine

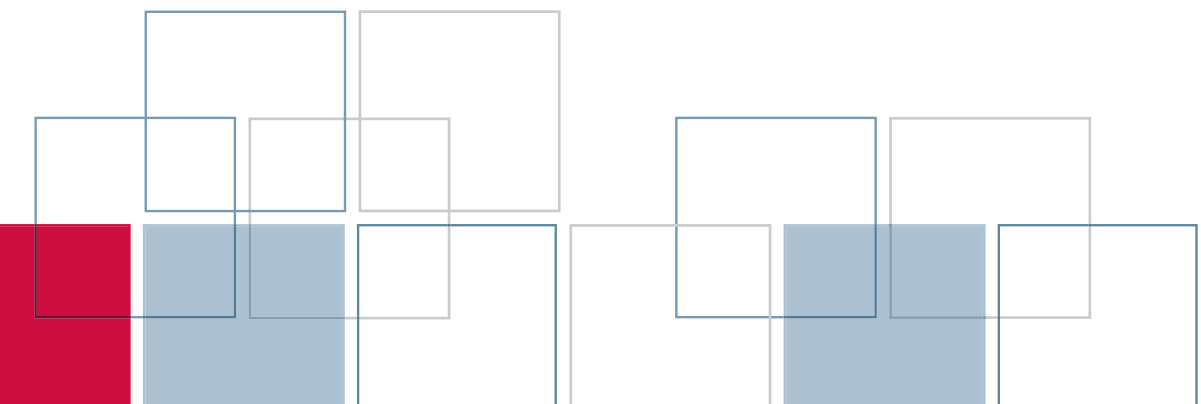
Evaluation

- Easily evaluate models using lift, gains, profit, and response graphs
  - Use a one-step process that shortens project time when evaluating multiple models
  - Define hit conditions and scoring expressions to interpret model performance
- Analyze overall model accuracy with coincidence matrices and other automatic evaluation tools

Deployment

Clementine offers a broad array of deployment capabilities to meet your organization's needs. Models built in Clementine can be directly deployed into other SPSS predictive applications and into other vendors' technologies.

- Clementine Solution Publisher (optional**)
  - Automate the export of all operations, including data access; data manipulations; model scoring, including combinations of models; and post-processing
  - Work in a runtime environment for executing image files on target platforms
  - Easily update solutions through small image files
- PredictiveCallCenter (optional**)
  - Automatically export full Clementine streams to the Integration Builder component of PredictiveCallCenter
  - Combine exported Clementine streams with PredictiveCallCenter models, business rules, and exclusions to optimize customer interactions
- Cleo™ (optional**)
  - A Web-based solution for rapid model deployment
  - Enable multiple users to simultaneously access and immediately score single records, multiple records, or an entire database, through a customizable browser-based interface
- Clementine Batch
  - Automate production tasks, while working outside the user interface
    - Automate Clementine processes from other applications or scheduling systems
    - Generate encoded passwords
    - Call Clementine processes via the command line
- Scripting
  - Use command-line scripts or scripts associated with Clementine streams to automate repetitive tasks in the user interface. Scripts generally perform the same types of actions that otherwise are carried out using a mouse or keyboard.
    - Execute selected lines from a stream, SuperNode, or stand-alone script, using an icon on the toolbar
    - Update stream parameters within a standalone script
- Export generated models as PMML
- Perform in-database scoring, which eliminates the need for—and costs associated with—transferring data or performing calculations on the client
  - Generate SQL from linear regression and factor-generated models
  - Generate SQL from decision trees and rule sets
- Use the bulk loading capabilities of your database
  - Increase performance during data export by using your database's bulk loader. Fine-tune various options, including row-wise or column-wise binding for loading via ODBC, and batch-size setting for batch commits to the database.



Scalability and integration

Scalability

- Use in-database mining to leverage parallel database implementations
- Minimize network traffic via intelligent field projection—Clementine pulls data only as needed from your data warehouse and passes only relevant results to the client

Integration

- ▣ Clementine supports the export of streams to SPSS' PredictiveCallCenter application through PredictiveCallCenter's Integration Builder component
- ▣ Export streams to the Integration Builder component with a new Predictive Applications Wizard
- ▣ In addition, Clementine integrates with:
  - ⊟ Microsoft Analysis Services, for building decision trees
  - ⊟ IBM DB2 Data Warehouse Enterprise Edition, for deploying models to Intelligent Miner Scoring and viewing Clementine models with Intelligent Miner Visualization
- ▣ On Sun Solaris systems, Clementine supports Kerberos security, a network authentication protocol
- Clementine integrates with SPSS for Windows and AnswerTree® and supports Cleo through a built-in Wizard
- It also integrates with other databases and market automation and software tools

System requirements

Clementine Client

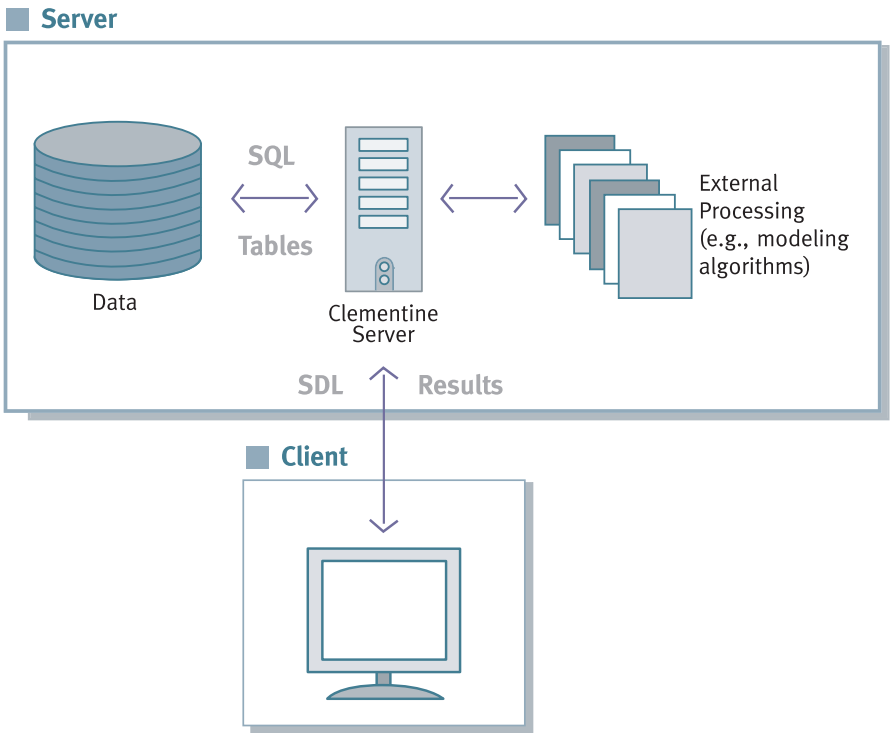
- Operating system: Windows XP Home Edition, Windows XP Professional, Windows 2000 Professional, Windows Server 2003, Windows 2000 Server, or Windows NT® 4.0 with Service Pack 6
- Memory: 512MB RAM recommended
- Minimum free disk space: 320MB
- A CD-ROM drive is required for installation
- Web browser: Internet Explorer version 5.x or later or Netscape® 6 or later, for running the online Help system

Clementine Server, Clementine Solution Publisher Runtime, and Clementine Batch

- Operating system: Windows Server 2003, Windows 2000 Server, Windows 2000 Professional, or Windows NT 4.0 with Service Pack 6 or later; Solaris 7, 8, or 9; HP-UX 11i; IBM AIX® 5.2; or OS/400® (on the IBM eServer iSeries™) V5R2 with OS/400 Portable Applications Solution Environment (PASE, 5722-SS1 Option 33)
- Hardware: Intel® Pentium®-compatible processor, if running on Windows; Sun UltraSPARC® II processor or better, if running on Solaris; PA-RISC processor and HP 9000 server or HP Workstation, if running on HP-UX; PowerPC® 233 MHz or faster and IBM RS/6000® for AIX; or IBM iSeries server for OS/400
- ▣ Hardware: 64-bit processor supported on SPARC servers running Solaris 9
- Memory: 512MB RAM minimum
- Minimum free drive space: 128MB of available disk space are required for installation. Additional free disk space is required to run the program (for temporary files). 1GB is recommended. For Clementine Solution Publisher Runtime, the minimum free disk space required to install the software is 64MB, plus at least twice the disk space of the amount of data to be processed.
- A network adapter running TCP/IP protocol
- A CD-ROM drive is required for installation

Clementine Application Templates:

- Clementine 7.2
- Memory: 512MB RAM recommended



Clementine uses a highly scalable, three-tiered architecture for data mining, as shown in this diagram. The Clementine Client tier (shown at the bottom of the illustration) passes stream description language (SDL) to Clementine Server. Clementine Server then analyzes particular tasks to determine which it can push to the database. After the database runs the tasks it can process, it passes only the relevant aggregated tables to Clementine Server. If you are using a CEMI, such as Text Mining for Clementine or Capri,™ Clementine Server passes the relevant tasks to that particular external process.

** Features are subject to change based on the final product release. ☐ Symbol indicates a new feature. **Separately priced modules*

Cleo

- Web server: Cleo requires at least one server computer that meets the following minimum requirements. Using additional processors, faster processors, and more RAM will improve performance.
  - Operating system: Windows 2000 Server or Windows Server 2003, Windows NT 4.0 Server with Service Pack 5 or higher (cannot be installed on Windows NT Terminal Server), or Solaris Version 7 or later
  - Hardware: Pentium-compatible processor, 500MHz or faster, if running on Windows; Sun UltraSPARC II or better, for Solaris
  - Memory: 512MB RAM
  - Minimum free disk space: 700MB
  - A graphics adapter with 800 x 600 resolution (SVGA) or higher, capable of displaying at least 256 colors
  - A network adapter running the TCP/IP protocol
- Repository: the system requires a database to serve as a repository for published content, framework settings, and other information. The following databases are supported:
  - Microsoft SQL Server 2000
  - Oracle 8i, version 8.1.7
- Data warehouse: the system can be configured to access data from a data warehouse or database. The system has only been tested with SQL Server 2000 and Oracle 8i databases.
- Web client: content is delivered to clients as standard HTML pages. Supported browsers include:
  - Internet Explorer version 5.5 with Service Pack 2 or version 6.0 for Windows
  - Internet Explorer version 5.2 for Macintosh®
  - Netscape 6.2

Text Mining for Clementine***

Client version requirements

- Clementine 7.1 or later
- Operating system: Windows XP Professional, Windows 2000 Professional, Windows NT 4.0 with Service Pack 6 or later, or Windows 2000 Server
- Memory: 512MB RAM
- Minimum free disk space: 85MB, plus space for databases
- Web Browser: Internet Explorer 5.0 or later or Netscape 6.0 or later is required to use the Viewer node

Server version requirements

- Operating system: Windows 2000 Server, Windows NT Server 4.0 with Service Pack 6 or later; Solaris 8 or 9
- Hardware: Pentium III processor 1GHz or faster, if running on Windows; Sun UltraSPARC II or better, if running on Solaris
  - Memory: 256MB RAM
  - Minimum free disk space: 85MB, plus space for databases

Web Mining for Clementine

Client version requirements

- Clementine 7.2 or later
- Operating system: Windows XP Home Edition, Windows XP Professional, Windows 2000 Professional with Service Pack 2 or later, Windows Server 2003 or Windows 2000 Server
- Memory: 512MB RAM
- Minimum free disk space: twice the amount of raw Web data being processed

Server version requirements

- Operating system: Windows XP Home Edition, Windows XP Professional, Windows 2000 Professional with Service Pack 2 or later, Windows Server 2003 or Windows 2000 Server
- Memory: 512MB RAM
- Minimum free disk space: twice the amount of raw Web data being processed

Capri for Clementine

Plug-in algorithm for detecting sequences

- Operating system: Windows 98, Windows 2000 Professional, Windows Server 2003, Windows 2000 Server, Windows NT 4.0 with Service Pack 3 or later, or Solaris 6
- Hardware: Pentium-compatible processor if running on Windows, or SPARC for Solaris
- Memory: 128MB RAM
- Minimum free drive space: 3MB
- A CD-ROM drive is required for installation